

# A Framework of 2D Fisher Discriminant Analysis: Application to Face Recognition with Small Number of Training Samples

Hui Kong Lei Wang Eam Khwang Teoh  
Nanyang Technological University  
Singapore 639798  
{pg03802060,elwang,eekteoh}@ntu.edu.sg

Jian-Gang Wang Ronda Venkateswarlu  
Institute for Infocomm Research  
Singapore 119613  
{jgwang,vronda}@i2r.a-star.edu.sg

## Abstract

*A novel framework called 2D Fisher Discriminant Analysis (2D-FDA) is proposed to deal with the Small Sample Size (SSS) problem in conventional One-Dimensional Linear Discriminant Analysis (1D-LDA). Different from the 1D-LDA based approaches, 2D-FDA is based on 2D image matrices rather than column vectors so the image matrix does not need to be transformed into a long vector before feature extraction. The advantage arising in this way is that the SSS problem does not exist anymore because the between-class and within-class scatter matrices constructed in 2D-FDA are both of full-rank. This framework contains unilateral and bilateral 2D-FDA. It is applied to face recognition where only few training images exist for each subject. Both the unilateral and bilateral 2D-FDA achieve excellent performance on two public databases: ORL database and Yale face database B.*

## 1 Introduction

When only  $t$  samples are available in an  $n$ -dimensional vector space with  $t < n$ , the sample covariance matrix  $\hat{\mathbf{C}}$  is calculated from the samples as

$$\hat{\mathbf{C}} = \frac{1}{t} \sum_{i=1}^t (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T \quad (1)$$

where  $\mathbf{m}$  is the mean of all the samples.  $(\mathbf{x}_i - \mathbf{m})^t$ 's are not linearly independent, because they are related by  $\sum_{i=1}^t (\mathbf{x}_i - \mathbf{m}) = 0$ . That is,  $\hat{\mathbf{C}}$  is a function of  $(t-1)$  or less linearly independent vectors. Therefore, the rank of  $\hat{\mathbf{C}}$  is  $(t-1)$  or less. This problem, which is called a Small Sample Size (SSS) problem [1], is often encountered in face recognition where  $t$  is very small and  $n$  is very large. In conventional LDA of face patterns, the criterion of measuring the discriminatory power of the projection vectors is to maximize the between-class scatter and meantime to minimize the within-class scatter of the projected samples. The optimal projection (transformation) can be readily computed

by solving a generalized eigenvalue problem. However, because of the SSS problem, the within-class covariance matrix,  $\mathbf{S}_w$ , is singular so that the numerical problem is introduced in solving the optimal discriminating directions.

To solve the SSS problem in 1D-LDA based face recognition, various schemes have been proposed so far. Swets and Weng's discriminant eigenfeatures [2], Belhumeur et al.'s Fisherface [3] and Zhao's discriminant component analysis [4] all used a two-stage PCA+LDA approach. Using PCA, the high dimensional face data are projected to a low dimensional space and then LDA is performed in this PCA space. However, the removed subspace may also contain some helpful information for recognition, and this removal may result in a loss of discriminative information. Chen et al. [5] suggested that the null space spanned by the eigenvectors of  $\mathbf{S}_w$  with zero eigenvalues contains the most discriminative information, hence an LDA method in the null space of  $\mathbf{S}_w$  was proposed, called N-LDA. However, as explained in [5], with the existence of noise, when the number of training samples is large, the null space of  $\mathbf{S}_w$  becomes small, and much discriminative information outside this null space will be lost. Another shortcoming is that this approach involves solving the eigenvalue problem for a very large matrix. A similar idea was proposed in [7] where the N-LDA is performed in the range space of  $\mathbf{S}_t$ .

Yu and Yang [6] proposed the Direct LDA (D-LDA) algorithm which also incorporates the concept of null space. It first removes the null space of the between-class scatter matrix,  $\mathbf{S}_b$ , and seeks a projection to minimize the trace of within-class covariance in the range space of  $\mathbf{S}_b$ . Because the rank of  $\mathbf{S}_b$  is smaller than that of  $\mathbf{S}_w$ , removing the null space of  $\mathbf{S}_b$  may lose part of or the entire null space of  $\mathbf{S}_w$ , which is very likely to be full-rank after the removing operation.

Wang and Tang [8] presented a random sampling LDA for face recognition with small number of training samples. This method can be regarded as a combination of weak classifiers. This paper concludes that both Fisherface and N-LDA encounter respective over-fitting problem for different reasons. A random subspace method and a random bag-

ging approach are proposed to solve them. A fusion rule is adopted to combine these random sampling based classifiers. A dual-space LDA approach [9] for face recognition was proposed to simultaneously apply discriminant analysis in the principal and null subspaces of the within-class covariance matrix. The two sets of discriminative features are then combined for recognition.

Recently, Two-Dimensional Principal Component Analysis (2D-PCA)[11] method has been discussed. 2D-PCA is based on 2D image matrices rather than column vectors as opposed to traditional 1D-PCA based approaches, thus the covariance matrix is quite small and can be evaluated more accurately in 2D-PCA than that in 1D-PCA. This makes the 2D-PCA achieve higher recognition rate than 1D-PCA. However, like 1D-PCA, 2D-PCA is only good at image representation rather than discrimination. When there are large pose- and illumination-variations in face images, the top eigenvectors in 2D-PCA do not model identity information but these external variations. It can be expected that 2D-PCA will be inferior to LDA based algorithms in such cases. Our experiments will also verify this conclusion in the later part of this paper.

To overcome the shortcoming in 2D-PCA and meanwhile to solve the SSS problem in 1D-LDA based algorithms, a novel 2D-FDA framework containing unilateral 2D-FDA and bilateral 2D-FDA is proposed. Similar to 2D-PCA, 2D-FDA constructs the between-class and within-class covariance matrices using the 2D image matrices, but different from 2D-PCA, the Fisher's criterion is adopted to find more discriminant information. In contrast to the  $\mathbf{S}_b$  and  $\mathbf{S}_w$  of 1D-LDA, the  $\mathbf{S}_b$  and  $\mathbf{S}_w$  obtained by 2D-FDA are not singular. As a result, the 2D-FDA has three important advantages over the 2D-PCA and 1D-LDA based algorithms. Firstly, the features are extracted using Fisher discriminant analysis instead of PCA, thus the discriminating ability is better than 2D-PCA. Secondly, it does not encounter SSS problem anymore when the training sample size is small. Thirdly, it takes full advantage of the discriminative information in the face space, and does not discard any subspace which may be valuable for recognition. In addition, within the framework of 2D-FDA, bilateral 2D-FDA is further developed. It shows better performance than its unilateral counterpart because incorporating bilateral projections extracts more discriminant information. Experiment results on ORL database and Yale face database B clearly demonstrate these advantages.

## 2 Unilateral 2D-FDA

Let  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_d]$  denotes an  $m \times d$  projection matrix, where  $\mathbf{w}_i$  is an  $m$ -dimensional column vector,  $i = 1, \dots, d$ . The idea is to project image  $\mathbf{X}$ , an  $m \times n$  matrix,

onto  $\mathbf{W}$  by the following linear transformation:

$$\mathbf{Y} = \mathbf{W}^T \mathbf{X} \quad (2)$$

Thus, we obtain a  $d \times n$  projected feature matrix  $\mathbf{Y}$  for each image. As in 1D-FDA, the discriminatory power of the projection vectors  $\mathbf{W}$  can be measured by the Fisher criterion [1], i.e., maximizing the between-class and meantime minimizing the within-class scatter of the projected samples. From this point of view, the Fisher criterion in 2D-FDA is adopted as follows:

$$J(\mathbf{W}) = \frac{\det(\mathbf{P}\mathbf{S}_b)}{\det(\mathbf{P}\mathbf{S}_w)} \quad (3)$$

where  $\mathbf{P}\mathbf{S}_b$  and  $\mathbf{P}\mathbf{S}_w$  are the between-class and within-class covariance matrices of the projected samples respectively,  $\det(\bullet)$  denotes the determinant of a matrix.

**Lemma 1** : Let  $\mathbf{S}_b, \mathbf{S}_w$  be the between-class and within-class covariance matrices of the original image matrix. Then,  $J(\mathbf{W}) = \frac{\det(\mathbf{W}^T \mathbf{S}_b \mathbf{W})}{\det(\mathbf{W}^T \mathbf{S}_w \mathbf{W})}$ .

**Proof** : Let  $\overline{\mathbf{M}}$  be the mean of all the training samples,  $\overline{\mathbf{M}}_i$  be the mean of the  $i$ -th class,  $\overline{\mathbf{M}}^p$  be the mean of all the projected samples,  $\overline{\mathbf{M}}_i^p$  be the mean of the  $i$ -th projected class.

Then,  $\mathbf{P}\mathbf{S}_b = \sum_{i=1}^L L_i (\overline{\mathbf{M}}_i^p - \overline{\mathbf{M}}^p) (\overline{\mathbf{M}}_i^p - \overline{\mathbf{M}}^p)^T = \sum_{i=1}^L L_i (\mathbf{W}^T \overline{\mathbf{M}}_i - \mathbf{W}^T \overline{\mathbf{M}}) (\mathbf{W}^T \overline{\mathbf{M}}_i - \mathbf{W}^T \overline{\mathbf{M}})^T = \sum_{i=1}^L L_i \mathbf{W}^T (\overline{\mathbf{M}}_i - \overline{\mathbf{M}}) (\overline{\mathbf{M}}_i - \overline{\mathbf{M}})^T \mathbf{W} = \mathbf{W}^T \mathbf{S}_b \mathbf{W}$ , where  $L$  is the total class number,  $L_i$  is the number of training samples in the  $i$ -th class,  $\mathbf{S}_b = \sum_{i=1}^L L_i (\overline{\mathbf{M}}_i - \overline{\mathbf{M}}) (\overline{\mathbf{M}}_i - \overline{\mathbf{M}})^T$ . Similarly, we can obtain  $\mathbf{P}\mathbf{S}_w = \mathbf{W}^T \mathbf{S}_w \mathbf{W}$ , where  $\mathbf{S}_w = \sum_{i=1}^L \sum_{j=1}^{L_i} (\mathbf{X}_i^j - \overline{\mathbf{M}}_i) (\mathbf{X}_i^j - \overline{\mathbf{M}}_i)^T$ .

Therefore, the Fisher criterion in Eq.3 can be converted to:

$$J(\mathbf{W}) = \frac{\det(\mathbf{W}^T \mathbf{S}_b \mathbf{W})}{\det(\mathbf{W}^T \mathbf{S}_w \mathbf{W})} \quad (4)$$

The vectors in  $\mathbf{W}$  that maximize Eq.4 are called the optimal discriminating projection axes.  $\square$

Since the above projection in Eq.2 is a unilateral left-multiplication, the 2D-FDA obtained in this way is called Unilateral 2D Fisher Discriminant Analysis (U2DFDA). For U2DFDA, we have the following theorem:

**Theorem 1**: The  $\mathbf{S}_w$  in U2DFDA is not singular.

**Proof** : Since  $\mathbf{S}_w = \sum_{i=1}^L \sum_{j=1}^{L_i} (\mathbf{X}_i^j - \overline{\mathbf{M}}_i) (\mathbf{X}_i^j - \overline{\mathbf{M}}_i)^T$ , another form of  $\mathbf{S}_w$  can be written as  $\mathbf{S}_w = \Phi_{S_w} \Phi_{S_w}^T$ , where  $\Phi_{S_w} = [\phi_1^{S_w}, \phi_2^{S_w}, \dots, \phi_L^{S_w}]$ ,  $\phi_i^{S_w} = [(\mathbf{X}_i^1 - \overline{\mathbf{M}}_i), (\mathbf{X}_i^2 - \overline{\mathbf{M}}_i), \dots, (\mathbf{X}_i^{L_i} - \overline{\mathbf{M}}_i)]$ ,  $i = 1, \dots, L$ ,  $j = 1, \dots, L_i$ .  $\mathbf{X}_i^j$  is the  $j$ -th training sample in the  $i$ -th class. The dimension of  $\Phi_{S_w}$  is  $m \times (n \sum_{i=1}^L L_i)$ , where  $m$  and  $n$  are the image's height and width. Since  $\text{rank}(\Phi_{S_w}^T \Phi_{S_w}) = \text{rank}(\Phi_{S_w} \Phi_{S_w}^T) = \text{rank}(\Phi_{S_w}^T) = \text{rank}(\Phi_{S_w})$  and

$rank(\Phi_{S_w}) = m^1$ , additionally, the dimension of  $\mathbf{S}_w$  is  $m \times m$ , we can conclude that  $\mathbf{S}_w$  is of full rank.  $\square$

In this way, the optimal projection vectors  $\mathbf{W}_{opt}$  can be obtained by directly solving the following generalized eigen-value problem.

$$\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{W} = \Lambda \mathbf{W} \quad (5)$$

where  $\Lambda$  is the diagonal matrix whose diagonal elements are eigenvalues of  $\mathbf{S}_w^{-1} \mathbf{S}_b$ .

The classification method is adopted as the one used in 2D-PCA [11]. In 1D-FDA based methods, the final dimension for classification is fixed to  $(C - 1)$ , where  $C$  is the number of classes. However, in 2D-FDA, the optimal number of *Fisher feature vector*,  $d$ , is not fixed. Since the  $\mathbf{S}_w$  is invertible,  $d$  can be at most equal to the image's height. However, the optimal  $d$  for classification is database-dependent, i.e., the optimal  $d$  is different for different databases. In our experiments, we will discuss the optimal dimensions for different databases.

### 3 Bilateral 2D-FDA

The above section describes a method of extracting the optimal discriminant directions via a left-multiplying U2DFDA. What if the projection is a right-multiplying operation? That is,

$$\mathbf{Y} = \mathbf{X} \mathbf{W} \quad (6)$$

In fact, it is trivial to check that the right-multiplying U2DFDA can be converted into left-multiplying U2DFDA by transposing the image matrix. Therefore, the right-multiplying *Fisher feature matrix*  $\mathbf{Y}_r = \mathbf{X} \mathbf{W}_r$  and  $\mathbf{W}_r$  can be obtained using Eq.5, where  $\mathbf{S}_b = \sum_{i=1}^L L_i (\overline{\mathbf{M}}_i - \overline{\mathbf{M}})^T (\overline{\mathbf{M}}_i - \overline{\mathbf{M}})$  and  $\mathbf{S}_w = \sum_{i=1}^L \sum_{j=1}^{L_i} (\mathbf{X}_i^j - \overline{\mathbf{M}}_i)^T (\mathbf{X}_i^j - \overline{\mathbf{M}}_i)$ .

Will the left-multiplying and right-multiplying U2DFDA achieve the same recognition rate or will they recognize the same batch of face images? Our experimental results show that sometimes they have the same recognition rate, but most of the time, their performance is different. The reason is that the calculations of  $\mathbf{S}_b$  and  $\mathbf{S}_w$  are different in the left-multiplying and right-multiplying U2DFDA. It can also be found that either in left-multiplying U2DFDA or right-multiplying U2DFDA, the calculations of  $\mathbf{S}_b$  and  $\mathbf{S}_w$  solely emphasize the dependency (correlation) among the row or column vectors of the image matrix and neglects the other one. Therefore, it may lose some information which

<sup>1</sup>It is known that  $rank(\Phi_{S_w}) \leq \min(m, n \sum_{i=1}^L L_i)$  and  $m \ll (n \sum_{i=1}^L L_i)$  in the area of visual pattern recognition, there is,  $rank(\Phi_{S_w}) \leq m$ . Further, we take an assumption that the rows of  $\Phi_{S_w}$  are independent of each other (the experiment results will demonstrate that this assumption can be well satisfied in the benchmark databases), it can be obtained that  $rank(\Phi_{S_w}) = m$ .

is helpful for discrimination. Considering this, a bilateral-projection scheme which is called Bilateral 2D Fisher Discriminant Analysis (B2DFDA) is proposed by combining  $\mathbf{Y}_l = \mathbf{W}_l^T \mathbf{X}$  and  $\mathbf{Y}_r = \mathbf{X} \mathbf{W}_r$ , where  $\mathbf{W}_l = [\mathbf{w}_1^l, \mathbf{w}_2^l, \dots, \mathbf{w}_{d_l}^l]$ ,  $\mathbf{W}_r = [\mathbf{w}_1^r, \mathbf{w}_2^r, \dots, \mathbf{w}_{d_r}^r]$  are the left- and right-multiplying optimal projection vectors respectively,  $d_l$  is the number of left-multiplying projection directions and it can be equal to the image's height at most.  $d_r$  is the number of right-multiplying projection directions and it can be equal to the image's width at most.

After performing the left- and right-multiplying U2DFDA,  $\mathbf{Y}_l$  and  $\mathbf{Y}_r$  are obtained for each image. They are combined together for recognition. The steps for recognition is as follows: firstly  $\mathbf{Y}_l$  and  $\mathbf{Y}_r$  are transformed into 1D vectors for each images, then PCA is applied onto these vectors. Finally, two shorter vectors can be obtained for each image and they are combined into one vector for classification.

### 4 The Essence of 2D-FDA

The covariance matrices in the U2DFDA appear to be physically meaningful in the matrix space rather than in the vector space. However, Theorem 2 will give another perspective to make the U2DFDA physically meaningful even in vector space, and will explicitly explain its essence. The similar explanation for 2DPCA has been made in [10].

**Theorem 2:** In the left-multiplying U2DFDA, the U2DFDA performed on the image matrices is essentially the conventional LDA method performed on the columns of the image matrices if each column is viewed as a computational unit.

**Proof :** Since  $\mathbf{S}_w = \sum_{i=1}^L \sum_{j=1}^{L_i} (\mathbf{X}_i^j - \overline{\mathbf{M}}_i) (\mathbf{X}_i^j - \overline{\mathbf{M}}_i)^T$ , another form of  $\mathbf{S}_w$  can be written as  $\mathbf{S}_w = \Phi_{S_w} \Phi_{S_w}^T$ .  $\Phi_{S_w} = [\phi_1^{S_w}, \phi_2^{S_w}, \dots, \phi_i^{S_w}]$ , and  $\phi_i^{S_w} = [(\mathbf{X}_i^1 - \overline{\mathbf{M}}_i), (\mathbf{X}_i^2 - \overline{\mathbf{M}}_i), \dots, (\mathbf{X}_i^{L_i} - \overline{\mathbf{M}}_i)] = [((\mathbf{X}_i^1(:, 1) - \overline{\mathbf{M}}_i(:, 1)), \dots, ((\mathbf{X}_i^1(:, n) - \overline{\mathbf{M}}_i(:, n))), \dots, ((\mathbf{X}_i^{L_i}(:, 1) - \overline{\mathbf{M}}_i(:, 1)), \dots, (\mathbf{X}_i^{L_i}(:, n) - \overline{\mathbf{M}}_i(:, n)))]$ , where  $\mathbf{X}_i^j$  is the  $j$ -th training sample in the  $i$ -th class, and  $\mathbf{A}(:, i)$  is the  $i$ -th column of matrix  $\mathbf{A}$ . Therefore,  $\mathbf{S}_w$  is constructed directly by the columns of the centered training image matrices. Similarly,  $\mathbf{S}_b$  is also constructed using the columns.

Therefore, the left-multiplying U2DFDA performed on the image matrices can be viewed as the conventional LDA performed on the columns of all the training samples if each column is viewed as a computational unit.  $\square$

**Theorem 3:** In the right-multiplying U2DFDA, the U2DFDA performed on the image matrices is essentially the conventional LDA method performed on the rows of the image matrices if each row is viewed as a computational unit.

The proof of Theorem 3 is similar to that of Theorem 2.

## 5 Experiment Results

The proposed U2DFDA and B2DFDA methods are tested on two commonly used face image databases, ORL and Yale face database B (YaleB). The ORL database is used to evaluate the performance of 2D-FDA under conditions where the pose, face expression and face scale vary. The YaleB is used to examine the performance of 2D-FDA when illumination varies significantly.

The ORL face database contains images from 40 individuals, each providing 10 different images. The facial expressions and facial details (glasses or no glasses) also vary. The images were taken with a tolerance for some tilting and rotation of the face of up to 20 degrees. Moreover, there is also some variation in the scale of up to about 10 percent. All images normalized to a resolution of  $46 \times 56$ .

From YaleB, altogether 640 images for 10 subjects are used (64 illumination conditions under the same frontal pose) in our experiments. All the images are preprocessed and normalized by translation, rotation, and scaling such that the two outer eye corners are in fixed positions. The image size is  $50 \times 40$ .

### 5.1 Random Grouping of Training and Testing Sets

To test the recognition performance with different training numbers,  $k$  ( $2 \leq k \leq 9$  for ORL database and  $2 \leq k \leq 12$  for YaleB) images of each subject are randomly selected for training and the remaining  $p-k$  ( $p$  is the total number of images for each subject, for ORL database,  $p$  equals 10; for YaleB,  $p$  is 64) images of each subject for testing. For each number  $k$ , 50 times of random selections are performed on ORL database and 100 times for YaleB. The final recognition rate is the average of all.

### 5.2 Experiment Setup

The first experiment is to investigate the performance of U2DFDA with different number of *Fisher feature vectors*. Without losing generality, the right-multiplying mode is used in U2DFDA. The maximum size of the *Fisher feature matrix* is  $56 \times 46$  for ORL database, i.e., containing at most 46 56-dimensional *Fisher feature vectors*; for YaleB, the maximum size of the *Fisher feature matrix* is  $50 \times 40$ , i.e., containing at most 40 50-dimensional *Fisher feature vectors*. We change the number of *Fisher feature vectors* from 1 to 46 for ORL database and from 1 to 40 for YaleB to see the effect on performance. It has been shown via theoretical analysis that there is no SSS problem in 2D-FDA. In order to verify this argument again through experiment, we focus on testing the performance of 2D-FDA when there are only few training samples for each subject, say, only 2, 3

or 4 training samples for each subject. Fig.1 (a) to (c) show the performance of U2DFDA on ORL database. The optimal number of the *Fisher feature vectors* in all the three trials is 3. Fig.1 (d) to (f) show the performance of U2DFDA on YaleB. The optimal number of the *Fisher feature vectors* in the three trials is 31, 27 and 23 respectively. From the experiment results, it can be seen that the optimal number of *Fisher feature vectors* for classification in U2DFDA is different on different database. Even on the same database, the optimal number will vary when the number of training samples for each subject is different.

The second experiment compares B2DFDA with U2DFDA. By fixing the optimal number of the *Fisher feature vectors* of the right-multiplying U2DFDA, we change the number of the *Fisher feature vectors* of the left-multiplying U2DFDA (from 1 to 56 for ORL database and from 1 to 50 for Yale face database B) and apply B2DFDA. Fig.1 (a) to (c) show the comparison of B2DFDA and U2DFDA on ORL database while Fig.1 (d) to (f) show the comparison results on Yale face database B. From this experiment, we find that B2DFDA can achieve higher recognition rate than U2DFDA, e.g., with an increase of up to 5 percentage on YaleB. Additionally, its performance variation with dimension change is smaller than that of U2DFDA.

The third experiment is to test the performance of U2DFDA and B2DFDA with different training numbers for each subject. A comparison is made between 2D-FDA and the state-of-the-art linear subspace methods. For Fisherface [3], the PCA subspace is constrained to  $(N - C)$ , the reserved dimension for classification is set to  $(C - 1)$ , where  $N$  is the total number of training samples,  $C$  is the number of classes. For D-LDA [6] and N-LDA [7], the dimension for classification are both set to  $(C - 1)$ . For 2D-PCA and U2DFDA, we select the optimal numbers of *Eigen feature vector* and *Fisher feature vector* for classification. The optimal numbers of 2D-PCA and U2DFDA for ORL database and YaleB are reported in (a) and (b) in Fig.2 respectively. For B2DFDA, the numbers of the right- and left-multiplying *Fisher feature vector* are set to be equal, both being the optimal number in U2DFDA. Two observations can be found from Fig.2. Firstly, on ORL, for 2D-PCA, with the increase of the number of training samples for each subject, the optimal number also increases. Contrarily, for U2DFDA, the optimal number decreases. On YaleB, for 2D-PCA, the optimal dimension is equally the maximum number of Eigen feature vectors, for U2DFDA, the optimal number drops as the number of training samples increases. Secondly, the optimal number for U2DFDA are smaller than that of 2D-PCA no matter what the number of training samples for each subject is. Hence, from the point view of storage-space requirement and computational load, U2DFDA is more efficient than 2D-PCA.

Fig.3 (a) shows the average recognition rate of all the

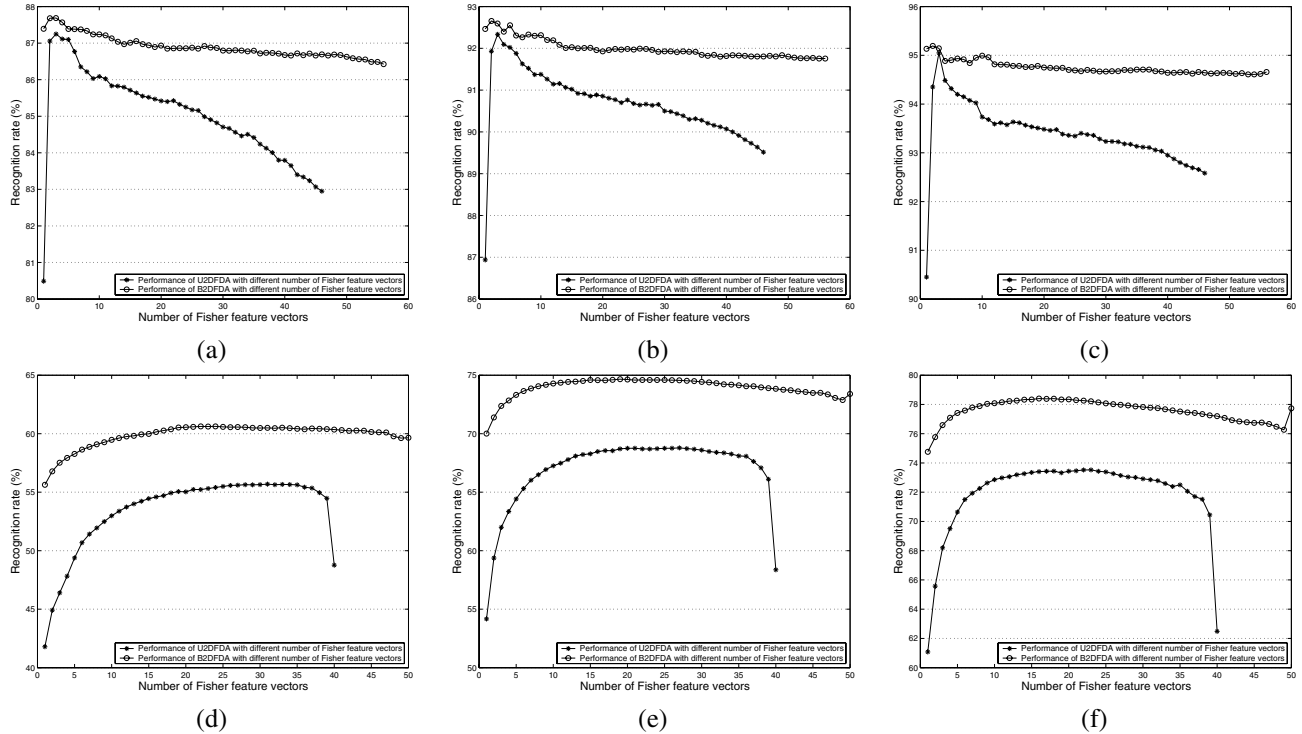


Figure 1: Comparison between U2DFDA and B2DFDA with different number of *Fisher feature vectors*. (a)-(c): Two, three and four training samples respectively for each subject on ORL database; (d)-(f): Two, three and four training samples respectively for each subject on YaleB.

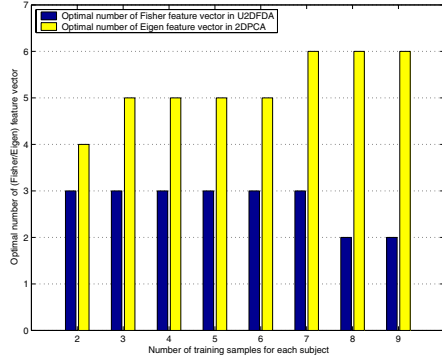
algorithms on ORL. It can be seen that the performance of U2DFDA and B2DFDA is much better than the other methods when the number of training samples is small. 2D-FDA outperforms the other linear LDA algorithms by up to 7 ~ 12 percentage, and surpasses the 2D-PCA by up to 3.5 percentage. We can also find that B2DFDA does outperform U2DFDA as expected by the theoretical analysis. 2D-PCA is superior to the linear LDA based algorithms on this database where there are no significant illumination variations. Fig.3 (b) shows the average recognition rate on YaleB. The performance of 2D-FDA is still much better than the other linear subspace methods when the number of training samples for each subject is small, however, with the increase of the number of training samples for each subject, the performance of Fisherface surpasses that of U2DFDA (when the training sample number is up to 7 for each subject). B2DFDA is much better than U2DFDA, with an increase of 2 ~ 5 percentage. However, Fisherface will outperform B2DFDA when the number of training samples for each subject increases up to a larger one (e.g., 15). This is because, in Fisherface method, the null space of  $\mathbf{S}_w$  is discarded, whose rank is  $C - 1$ . When the number of training samples for each subject is small (e.g., 2), the rank of null space of  $\mathbf{S}_w$  is comparable to the rank of range space

of  $\mathbf{S}_w$ . Therefore, discarding the whole null space of  $\mathbf{S}_w$  is equivalent to losing a large quantity of discriminant information. However, with the number of training samples increasing, the rank of null space of  $\mathbf{S}_w$  is much smaller than the rank of range space of  $\mathbf{S}_w$  and discarding the whole null space of  $\mathbf{S}_w$  will lose relatively little useful information.

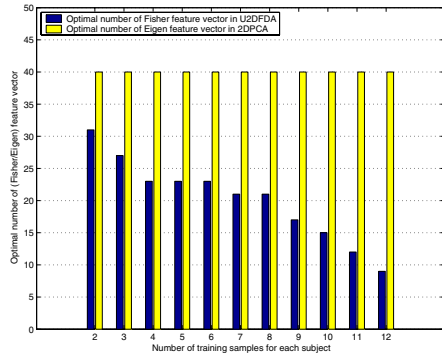
We also find that the performance of 2D-PCA is not so good as that on ORL database. This is consistent with our previous analysis that 2D-PCA is a good pattern representation methods but not a good discriminant one. When the images are taken under large illumination variations, its top eigenvectors mainly model the external illumination variation but not the identity information. This is a commonly known drawback of PCA and 2D-PCA for face recognition with significant illumination changes.

## 6 Conclusion

A 2D-FDA framework containing U2DFDA and B2DFDA is proposed to solve the SSS problem in face recognition. The key advantage of the 2D-FDA over the existing 1D-LDA based algorithms is that the within-class scatter matrix is generally not singular, which leads to the direct utilization



(a)



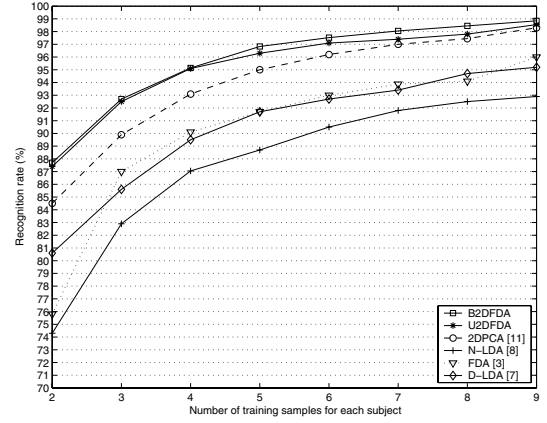
(b)

Figure 2: (a) and (b) show the optimal number of *Fisher feature vector* and *Eigen feature vector* in U2DFDA and 2D-PCA with different number of training samples for each subject on the ORL and YaleB

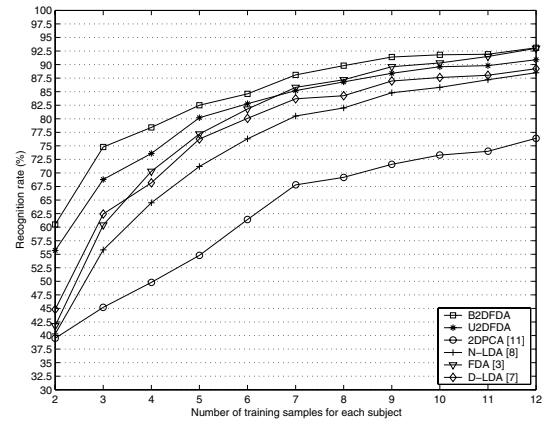
of the Fisher criterion for optimal classification. Extensive experimental work shows that the 2D-FDA framework outperforms the current linear subspace methods.

## References

- [1] K. Fukunaga, "Introduction to Statistical Pattern Recognition," Academic Press, second edition, 1991.
- [2] D.L. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," IEEE Trans. on PAMI, 1996.
- [3] P.N. Belhumeur, J. Hespanha and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," IEEE Trans. on PAMI, 1997.
- [4] W. Zhao, "Discriminant Component Analysis for Face Recognition," In Proc. International Conference on Pattern Recognition, 2000.
- [5] L. Chen, H. Liao, M. Ko, J. Lin and G. Yu, "A New LDA-based Face Recognition System Which can Solve the Small Sample Size Problem," Pattern Recognition, 2000.



(a)



(b)

Figure 3: (a) and (b) show the recognition rate of U2DFDA and B2DFDA compared with the other linear subspace methods on the ORL and Yale face database B

- [6] H. Yu and J. Yang, "A direct lda algorithm for high-dimensional data with application to face recognition," Pattern Recognition, 2001.
- [7] R. Huang, Q.S. Liu, H.Q. Lu and S.D. Ma, "Solving the Small Sample Size Problem of LDA," Proc. IEEE Conf. ICPR, 2002.
- [8] X. Wang and X. Tang, "Random sampling LDA for face recognition," Proc. IEEE Conf. CVPR, 2004.
- [9] X. Wang and X. Tang, "Dual-space linear discriminant analysis for face recognition," IEEE Conf. CVPR, June 2004.
- [10] H. Kong, L. Wang, E.K. Teoh, J.G. Wang and V. Ronda, "Generalized 2D Principal Component Analysis," to appear in IEEE Conf. IJCNN, Canada, 2005.
- [11] J. Yang, D. Zhang, A.F. Frangi and J. Yang, "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," IEEE Trans. on PAMI, 2004.